

RELIABILITY FOR INTERCONNECT FABRICS

This is a continuation-in-part of U.S. Application No. 09/707,227, filed
5 November 16, 2000, the contents of which are hereby incorporated by
reference.

Field of the Invention:

The present invention relates to the field of networks. More particularly,
10 this invention relates to reliability of networks.

Background of the Invention:

An interconnect fabric provides for communication among a set of
nodes in a network. Communications originate within the network at a source
15 node and terminate at a terminal node. Thus, a wide variety of networks may
be viewed as a set of source nodes that communicate with a set of terminal
nodes via an interconnect fabric. For example, a storage area network may be
arranged as a set of computers as source nodes which are connected to a set of
storage devices as terminal nodes via an interconnect fabric that includes
20 communication links and devices such as hubs, routers, switches, etc. Devices
such as hubs, routers, switches, etc., are hereinafter referred to as interconnect
devices. Depending on the circumstances, a node may assume the role of
source node with respect to some communications and of terminal node for
other communications.

25 The communication requirements of an interconnect fabric may be
characterized in terms of a set of flow requirements. A typical set of flow
requirements specifies the required communication bandwidth from each
source node to each terminal node. The design of an interconnect fabric
usually involves selecting the appropriate arrangement of physical
30 communication links and interconnect devices and related components that will
meet the flow requirements.

An interconnect fabric that meets the minimum flow requirements under ideal conditions will not necessarily meet the flow requirements under other conditions, such as in the event of a failure of a communication link, interconnect device or related component. Therefore, network designers typically address these reliability considerations by building in excess capacity or redundancy to help meet flow requirements under adverse conditions. Prior techniques are largely ad hoc and, thus, tend to be time-consuming, error-prone and may result in an over-provisioned interconnect fabric.

Summary of the Invention:

A technique is disclosed for providing reliability to an interconnect fabric for communication among a set of nodes. The technique may be used to efficiently and programmatically produce a cost-effective interconnect fabric having a degree of reliability over a range of design problems.

In one aspect, reliability is provided to an interconnect fabric for communication among a set of nodes. Ports associated with each node are partitioned into a first set of ports and a second set of ports. A primary interconnect fabric is formed among the first set of ports in response to a set of flow requirements. A backup interconnect fabric is formed among the second set of ports. The backup interconnect fabric carries a portion of communications carried by the primary fabric so as to protect against occurrence of a failure in the primary fabric.

In another aspect, reliability is provided to an interconnect fabric for communication among a set of nodes. One or more failure modes are identified in a primary interconnect fabric that carries communications among the set of nodes via a first set of ports of the nodes. A backup interconnect fabric is formed among a second set of ports of the nodes for carrying a portion of the communications of the primary fabric so as to protect against occurrence of any single one of the failure modes of the primary fabric.

In yet another aspect, reliability is provided to a design for an interconnect fabric for communication between a set of nodes. A set of design

information includes a set of flow requirements for the interconnect fabric. A fabric design tool generates a primary design for the interconnect fabric among of first set of ports of the nodes. The primary design is in response to the flow requirements. The design tool also generates a backup design for the

5 interconnect fabric among a second set of ports for the nodes. The backup design carries a portion of communications carried by the primary fabric so as to protect against occurrence of any single one of failure modes of the primary fabric.

The backup interconnect fabric may be formed by generating

10 arrangements of flow sets in response to the flow requirements, determining feasibility of merging pairs of candidate flow sets and merging a pair of the flow sets. The feasibility of merging candidate flow sets may depend on a sum of flow requirements in the candidate flow sets that are interrupted by a single failure in the primary fabric or may depend on a highest sum of flow

15 requirements in the candidate flow sets that are interrupted by different failures in the primary interconnect fabric.

Brief Description of the Drawings:

The present invention is described with respect to particular exemplary

20 embodiments thereof and reference is accordingly made to the drawings in which:

Figure 1 shows a method for providing reliability to an interconnect fabric according to an embodiment of the present invention;

Figure 2 shows an arrangement of flow sets in an interconnect fabric for

25 an example design according to an embodiment of the present invention;

Figure 3 shows how ports at each node may be partitioned into sets for the example design according to an embodiment of the present invention;

Figure 4 shows a method for forming an interconnect fabric among a first set of ports according to an embodiment of the present invention;

30 Figures 5-6 show a primary interconnect fabric for the example design evolving according to an embodiment of the present invention;

Figure 7 shows a method for forming a backup interconnect fabric among a second set of ports according to an embodiment of the present invention;

Figures 8-11 show a backup interconnect fabric for the example design evolving according to an embodiment of the present invention;

Figure 12 shows primary and backup interconnect fabrics for the example design according to an embodiment of the present invention; and

Figure 13 shows a fabric design tool that may employ techniques of the present invention to provide reliability to an interconnect fabric design.

Detailed Description of a Preferred Embodiment:

Figure 1 shows a method 100 for providing reliability to an interconnect fabric according to an embodiment of the present invention. The method 100 partitions ports at each node into sets and forms primary and backup interconnect fabrics among the ports of each set based on flow requirements among the nodes. Reliability is provided because multiple fabrics interconnect the nodes. The backup interconnect fabric can preferably carry at least a portion of the communications of the primary fabric so as to protect against a failure of an element in the primary fabric. Preferably, for every failure in the primary fabric, the backup fabric is able to route the affected flows. For example, if a link breaks in the primary network, the backup fabric should be able to handle the flows previously routed through the broken link.

In a step 102, a set of nodes to be interconnected by an interconnect fabric, and flow requirements among the nodes, are determined. Table 1 shows an example set of flow requirements for an interconnect fabric under design.

	Terminal Node 50	Terminal Node 52	Terminal Node 54
Source Node 40	a	b	c
Source Node 42	d	e	f

The flow requirements in this example specify two source nodes (source nodes 40-42 in the figures below) and three terminal nodes (terminal nodes 50-54 in the figures below). If an interconnect fabric is to meet the flow requirements, it must contain communication paths between all pairs of the source and terminal nodes 40-42 and 50-54 having positive flow requirements and must have sufficient bandwidth to support all of the flow requirements simultaneously.

In one embodiment, the source nodes 40-42 are host computers and terminal nodes 50-54 are storage devices. Thus, the interconnect fabric under design may be storage area network.

The bandwidth values for flows a, b, c, d, e and f may be numbers expressed in units of megabits per second (Mb/s). For this example, assume that each of the flows a, b, c, d, e and f have a bandwidth requirement of 33 Mb/s.

In other embodiments, there may be multiple flow requirements between a given source and terminal node pair. In such embodiments, the cells of Table 1 would contain a list of two or more entries. And, depending on the circumstances, a node may assume the role of source node with respect to some communications and of terminal node for other communications.

Figure 2 shows an initial arrangement of flows for the flow requirements obtained at step 102 for this example. Each entry in the flow requirements table is represented by a communication path or flow between pairs of nodes.

More particularly, flow a is between the source node 40 and terminal node 50, flow b is between source node 40 and terminal node 52, flow c is between source node 40 and terminal node 54, flow d is between source node 42 and terminal node 50, flow e is between source node 42 and terminal node 52, flow f is between source node 42 and terminal node 54.

At step 104, the ports of each node may be partitioned into sets. For example, the ports at each node may be divided into two sets. In one embodiment, a first set includes all of the ports for each node, save one, and a second set includes the remaining port not assigned to the first set. In other embodiments, the ports of each node could be further divided into an additional number of (k) sets. In which case, additional fabrics may be used to interconnect the additional sets of ports to provide even greater redundancy and reliability. Figure 3 shows how ports at each node may be partitioned into two sets for the example design. In the example, each of nodes 40 and 42 has two ports and each of nodes 50, 52 and 54 has four ports. Assume for this example that each of these ports has a bandwidth capacity of 100 Mb/s. Note that if a node has only one port, the port may be split among the sets by connecting an interconnect device having at least three ports, such as a hub or repeater, to the port. The remaining ports of the interconnect device may then be partitioned into the sets as ports belonging to the node.

In the example, a first set of ports includes one port of each of the nodes 40 and 42 and three ports of each of the nodes 50, 52 and 54. A second set of ports includes one port of each of the nodes 40, 42, 50, 52 and 54. The first set includes those ports to the left of a dotted line (shown in Figure 3) which divides each node, while the second set includes those of ports shown to the right of the dotted line.

In a step 106 (Figure 1), a primary interconnect fabric is formed among a first set of ports to support the flow requirements for the design. Thus, the flows of Figure 2 are shown in Figure 3 as being supported by the first set of ports for each node.

Figure 4 shows a method 200 for forming an interconnect fabric among a first set of ports according to an embodiment of the present invention. The method 200 is disclosed in U.S. Application No. 09/707,227, filed November 16, 2000, the contents of which are hereby incorporated by reference, and may
5 be performed during the step 106 of Figure 1. It will be apparent, however, that other techniques for forming an interconnect fabric, such as manual or other methods, may be used in the step 106.

The method 200 partitions the flow requirements of the interconnect fabric into flow sets and iteratively merges the flow sets while taking into
10 account the feasibility and cost of implementing the interconnect fabric.

At step 202, an arrangement of flow sets in the interconnect fabric is determined in response to the set of flow requirements for the source and terminal nodes. In one embodiment, step 202 is performed by generating a flow set for each flow specified in the flow requirements for the interconnect
15 fabric. Thus, each of flows a, b, c, d, e and f of the example is initially included in a corresponding flow set having one flow.

At step 204, port violations which are associated with the arrangement of flow sets among the first set of ports are determined. In the example, port violations are determined for the first set of ports for each source node 40-42
20 and each terminal node 50-54. In general, the number of port violations is equal to the sum, over all flow sets, of the number of required physical communication links to the node from that flow set, minus the number of available ports in the set of ports at the node. Each flow set may require one or more physical communication links to a given source or terminal node in the
25 network.

In this example (Figure 3), the number of port violations for a node is equal to the number of flow sets connected to the node minus the number of available ports in first set of ports for the node because each flow set is carried by one physical communication link in the interconnect fabric. Thus, the
30 source node 40 has a port violation of two since each of its three flow sets requires one physical communication link to the source node 40 and the source

node 40 has only one available port in the first set. Similarly, the source node 42 also has a port violation of two since each of its three flow sets requires one physical communication link to the source node 42 and the source node 42 has only one available port in the first set. The terminal nodes 50-54 have no port violations since the number of ports in the first set is greater than the number of flow sets. More particularly, each of the terminal nodes 50-54 has three available ports in the first set, but only two flow sets that each require one physical communication link to the node.

In other examples, the number of available ports in the first set for the source nodes 40-42 and the terminal nodes 50-54 may differ and the number of physical communication links required by a flow set on a given source or terminal node it connects to may exceed one.

At step 206 (Figure 4), at least one of the port violations is alleviated by merging a pair of the flow sets. Step 206 initially involves selecting the pair of flow sets in the current interconnect fabric that are to be merged. Initially, a candidate pair of flow sets is chosen that would alleviate the port violation on a node with the greatest port violation if merged. If there is more than one such candidate pair then one of the candidate pairs that alleviates a port violation on a node having the next greatest port violation is chosen from among them. If there is more than one such candidate pair then a pair of them that would be least costly to merge is chosen. The cost of merging two candidate pairs may be determined by choosing the least expensive interconnect device that is feasible for the merged flow set.

In the current state of the example interconnect fabric shown in Figure 3, the source nodes 40 and 42 each has a port violation of two, which is the worst port violation in the network, and its corresponding flow sets are candidates for merging at step 206. For example, the pair of flow sets having flows a and b or the pair of flow sets having flows b and c or the pair of flow sets having flows a and c may be chosen as candidate pairs. All of these candidate pairs, if merged, would alleviate one port violation from the terminal node 40 but none of them would alleviate any other port violations. Therefore, the cost of

merging the candidate pairs may be used to select the candidate pair of flow sets at step 206. For example, the communication link and/or interconnect device and/or ports that are used to merge the flow sets having flows a and b may be less expensive than the equivalent components needed to merge the other equally qualified candidate pairs. For example, assuming that link cost is related to length, replacing two links over a longer distance with a single link would likely reduce costs more than if over a shorter distance.

The candidate pairs of flow sets considered at step 206 must be feasible to merge. An example of a pair of flow sets that is not feasible to merge is a pair for which an interconnect device of sufficient bandwidth is not available. For example, a flow set having 60 units of bandwidth cannot be merged with a flow set having 50 units of bandwidth if the highest bandwidth interconnect device available has 100 units. Another example of a pair of flow sets that is not feasible to merge is a pair that would exceed the available ports on every available interconnect device of the resulting flow set. Candidate pairs that are not feasible to merge are bypassed at step 206 in favor of other candidate pairs.

If port violations still exist in the interconnect fabric after step 206, then another candidate pair of flow sets is selected and merged in a repeat of step 206. The method 200 loops through steps 204-206 in an iterative fashion until all port violations are eliminated or until no further merges are feasible.

Figure 5 shows an interconnect fabric including a flow set having an interconnect device 160, a flow of a and a flow of b that is the result of the merger of the flow set having the flow of a and the flow set having the flow of b. Assume that the interconnect device 160 and other available devices are hubs that each have an aggregate bandwidth capacity of 100 Mb/s. This means that the total bandwidth of flows entering or exiting the device cannot exceed 100 Mb/s. The merger of flows a and b by the device 160 is feasible insofar as the aggregate of these flows is 66 Mb/s, which is less than the maximum bandwidth for the device 160, which in the example is 100 Mb/s.

In a next pass through the step 206, a pair of flow sets from among those having the flows d, e and f may be merged to alleviate the port violation of two

at the source node 42. Thus, as shown in Figure 5, the interconnect fabric includes a flow set having an interconnect device 162, a flow of e and a flow of f that is the result of the merger of the flow set having the flow of e with the flow set having the flow of f. Thus, Figure 5 shows the interconnect fabric after two passes through the step 206. The merger of flows e and f by the device 162 is feasible insofar as the aggregate of these flows is 66 Mb/s, which is less than the maximum bandwidth for the device 162, which in the example is 100 Mb/s.

At this point, the interconnect fabric has a port violation of one at each of the source nodes 40 and 42. Then, another pass through the step 206 may result in the selection and merger of the flow set corresponding to the interconnect device 160 and with the flow set including the flows a and b with the flow set including the flow c which alleviates the port violation of the source node 40. The merger of flows a, b and c by the device 160 is feasible insofar as the aggregate of these flows is 99 Mb/s, which is less than the maximum bandwidth for the device 160, which in the example is 100 Mb/s.

A further pass through the step 206 may result in the selection and merger of the flow set corresponding to the interconnect device 162 and with the flow set including the flows e and f with the flow set including the flow d which alleviates the port violation of the source node 42. The merger of flows d, e and f by the device 160 is feasible insofar as the aggregate of these flows is 99 Mb/s, which is less than the maximum bandwidth for the device 160, which in the example is 100 Mb/s.

Figure 6 shows an interconnect fabric that results from these additional passes through step 206. At this point, the interconnect fabric has no port violation remaining. Accordingly, Figure 6 shows a primary interconnect fabric that interconnects the first set of ports for each node and that will support the flow requirements for the network. Note that if cost savings may be obtained by merging additional flow sets, such mergers may also be accomplished if feasible by making another pass through step 206.

Returning to the method 100 of Figure 1, once the primary interconnect fabric has been formed in the step 106 among the first set of ports, a secondary or backup interconnect fabric may be formed in the step 108 among the second set of ports. This backup interconnect fabric carries at least a portion of the flows that the primary fabric carries in order to protect against a failure of an element in the primary fabric. While the primary fabric contains communication paths having sufficient bandwidth to support all of the flow requirements simultaneously, the secondary fabric need not support all of the flow requirements simultaneously.

A method 300 illustrated in Figure 7 may be employed in the step 108 of Figure 1 to form the backup interconnect fabric. At step 302, possible failure modes of the primary fabric may be analyzed to determine which failure modes are dominant. A failure mode is dominant over other failure modes that would be masked by the occurrence of the dominant failure. In other words, a set of flows whose routes are interrupted by a dominated failure is a subset of the flows affected by a dominating failure. For example, a possible failure mode may be the failure of a single port of an interconnect device such that communication via that port ceases, but other ports of the interconnect device remain functional. Another failure mode may be the failure of the entire interconnect device such that all communication via the device ceases. The failure of the entire interconnect device can be said to dominate the failure of one of its ports since the occurrence of the dominant failure would mask the failure of the port. As another example, a possible failure mode may be the failure of a link which connects a flow between a source and a terminal node without passing through an interconnect device. Such a failure would be considered dominant since it would mask a failure of either the port at the source node that is connected to the link or the port at the terminal node that is connected to the link.

While the complete failure of a source node or terminal node could be said to dominate other failures, the backup interconnect fabric is preferably primarily protective of communications between the nodes, rather than of the

nodes themselves. Accordingly, such possible failures of source or terminal nodes may be disregarded for purposes of step 302.

Thus, in one embodiment, the dominating failures identified in the step 302 include interconnect devices in the primary fabric and links in the primary fabric that connect flows directly between source and terminal nodes (without the flows passing through any interconnect devices).

In the example, the interconnect devices 160 and 162 are identified in step 302 as dominant failure points. Because there are no links that connect flows directly between source and terminal nodes in the example, no such links are identified as dominant failures for the example.

In a step 304, each flow that is associated with each dominant failure may be identified. In other words, for each dominant failure identified in step 304, each flow that would be interrupted in the event of the dominant failure may be identified. The flows may also be grouped according to the dominant failure. Thus, for each dominant failure, the affected flows may be grouped together.

In the example, for the dominating failure of the interconnect device 160, the flows that would be interrupted include the flow of a, the flow of b and the flow of c. Thus, the group of flows associated with the failure of device 160 includes the flows a, b and c. For the dominant failure of the interconnect device 162, the flows that would be interrupted include the flow of d, the flow of e and the flow of f. Thus, the group of flows associated with the failure of the device 162 includes the flows d, e and f.

In a step 306, port violations which are associated with the arrangement of flow sets among the second set of ports are determined. The arrangement of flow sets may be determined from the step 202 (Figure 4) and, thus, the step 202 need not be repeated for the backup fabric. However, the port violations for step 306 may differ from those of step 204 since the second set of ports for the backup fabric will likely include different numbers of available ports compared to the primary fabric.

Figure 8 shows an initial arrangement of flow sets for the backup fabric for the example. The flow sets are similar to those shown in Figure 3 except that they now interconnect the second set of ports. In the example, port violations are determined for the second set of ports for each source node 40-42 and each terminal node 50-54. The source nodes 40-42 each have a port violation of two since each of their three flow sets requires one physical communication link to the node and each has only one available port in the second set. The terminal nodes 50-54 each have a port violation of one since the number of ports in the second set is one less than the number of flow sets connected to the nodes.

In a step 308, feasibility of possible merges is determined. Recall that that primary fabric is designed to accommodate all of the flows simultaneously. The backup fabric, however, need only accommodate a portion of the flows at any one time. Feasibility of possible merges for the backup fabric in step 308 is determined based on these more limited considerations. In one embodiment, the backup fabric need only provide for flows that are interrupted by the occurrence in the primary fabric of a single dominant failure. Thus, where only one of two different flows would be interrupted during the occurrence of a single dominant failure, their bandwidth requirements need not be simultaneously met. Rather, the worst-case bandwidth requirement for merging the two flows in the backup fabric is the greater requirement of the two flows. For example, if one such flow requires 50 units of bandwidth and the other flow requires 60 units of bandwidth, the worst-case bandwidth requirement is 60 units of bandwidth. However, for pairs of flows that would both be interrupted by the occurrence of a dominant failure, their bandwidth requirements would need to be met simultaneously in order for the network to be able to withstand such a failure. For example, if one such flow requires 50 units of bandwidth and the other requires 60 units, then the aggregated bandwidth requirement to be met by the backup fabric in the event of the failure is 110 units of bandwidth.

Where multiple sets of flows would each be interrupted by the occurrence of different failures, then the worst-case is the greatest sum of flow requirements among the sets. For example, assume that two flows that require 50 and 60 units of bandwidth, respectively, would be interrupted by one failure.

5 Assume also that three other flows each requiring 30 units of bandwidth each would be interrupted by a different failure. The sum of flow requirements for the first set is 110 (50 added to 60), whereas the sum of flow requirements for the second set is 90 (three times 30). Accordingly, the worst-case that needs to be considered for merging these flow sets is the highest sum of 110 units of
10 bandwidth.

Initially, for the example backup fabric of Figure 8, possible merges include the merger of flows connected to any of the nodes 40-42 or 50-54. This is because each has a port violation of at least one. Further, a merger of any pair of the flows at any node is feasible insofar as the combined bandwidth
15 in each case would be 66 Mb/s which is less than the maximum bandwidth of devices of 100 Mb/s.

Then, in a step 310, at least one of the port violations is alleviated by merging a pair of the flow sets. Because the source nodes 40 and 42 have the worst port violations, a pair of flow sets at the node 40 may be merged first.

20 For example, the flow sets having flows a and b may be merged by an interconnect device 164, as shown in Figure 9. The method 300 continues to loop through steps 306-310 in an iterative fashion until all port violations are eliminated or until no further merges are feasible. Thus, in a next pass through the steps 306-310, a pair of flow sets at the node 40 may be merged. For
25 example, the flow sets having flows e and f may be merged by an interconnect device 166, as is also shown in Figure 9. Thus, Figure 9 shows the backup fabric after these two merges have been completed.

Referring to Figure 9, each of the nodes 40-42 and 50-54 now has a port violation of one since each node has one available port in the second set and
30 the two flow sets connected to each node each require one physical communication link. In a next pass through the steps 306-310, the flow set

having the flow of c may be merged with the flow set having the flow of a, the flow of b and the interconnect device 164. While each of the flows a, b and c would be interrupted by the failure of the device 160 (Figure 6), merging them is feasible in the backup fabric insofar as their aggregated bandwidth

5 requirement of 99 Mb/s is less than the maximum bandwidth capacity of the interconnect device 164 of 100 Mb/s. Similarly, in a next pass through the steps 306-310, the flow set having the flow of d may be merged with the flow set having the flow of e, the flow of f and the interconnect device 166. While each of the flows e, f and g would be interrupted by the failure of the device
10 162 (Figure 6), merging them in the backup fabric is feasible insofar as their aggregated bandwidth requirement of 99 Mb/s is less than the maximum bandwidth capacity of the interconnect device 164 of 100 Mb/s. Figure 10 shows the backup interconnect fabric after these two merges have been completed.

15 Recall that for the primary fabric of Figure 6, no additional merges were feasible since any such merges would exceed the maximum bandwidth capacity of the interconnect devices. As mentioned, however, for the backup interconnect fabric, not all bandwidth requirements are additive when determining whether a merge is feasible. Thus, in the example, the bandwidth
20 requirements of the group of flows including the flows a, b and c are additive with respect to each other because they are all associated with the dominant failure of the interconnect device 160 (Figure 6). Also in the example, the bandwidth requirements of the group of flows including the flows d, e and f are additive with respect to each other because they are all associated with the
25 dominant failure of the interconnect device 164 (Figure 6). However, the bandwidth requirements for these flows are not summed with flows in another group. Rather, only the highest bandwidth requirement need be accommodated in order for the network to recover from a dominant failure.

Thus, in the example, to determine the feasibility of merging the flow
30 set having the flows a, b and c with the flow set having the flows d, e and f only the greater bandwidth of the two flow sets needs to be accommodated by

an interconnect device. Because each of these flow sets has a bandwidth requirement of 99 Mb/s, the greater of the two is also 99 Mb/s. Because this requirement is less than the maximum bandwidth capacity of the available interconnect devices, this means that these two flow sets can be merged to alleviate the port violations remaining at the terminal nodes 50-54. This is shown in Figure 10, where the device 164 merges the flow set having the flow of a, the flow of b and the flow of c with the flow set having the flow of d, the flow of e and the flow of f. In addition, the device 166 is eliminated.

The backup fabric of Figure 11 includes fewer interconnect devices than the primary interconnect fabric of Figure 6. Thus, the backup fabric can be implemented cost effectively, and with fewer ports at the source and terminal nodes, while still providing for recovery of the network in the event any single dominant failure occurs.

Note that in Figures 6 and 11, each flow set has at most one interconnect device associated with it and all of the flows for the flow set are routed through that interconnect device or a single communication link, if possible. Thus, the methods 200 and generally result in fabric designs of a single layer in which there are no links between device nodes.

Under certain circumstances, a single-layer fabric may not eliminate all of the port violations. In which case, the methods 200 and 300, by themselves may not result in a fabric design in which there are no port violations. Thus, in one embodiment, the present invention may address remaining port violations by recursively generating one or more additional layers of interconnect fabric nodes. For port violations at source nodes, the problem (i.e. the current fabric configuration and the applicable design information) may be recast such that the device nodes are treated as the terminal nodes. Then, one or more additional layers of device nodes may be inserted between the source nodes and the device nodes to relieve the port violations at source nodes. This results in links between device nodes and, thus, increases the number of layers in the interconnect fabric. Similarly, for terminal port violations, the problem may be recast such that the device nodes are treated as the source nodes. Then, one or

more additional layers of device nodes may be inserted in between the device nodes and the terminal nodes to relieve the terminal node port violations. This also results in links between the device nodes and, thus, increases the number of layers in the interconnect fabric. Such a technique is disclosed in co-
5 pending U.S. Application No. _____, entitled, "Designing Interconnect Fabrics," and filed December 19, 2001, the contents of which are hereby incorporated by reference and which is continuation-in-part of U.S. Application No. 09/707,227, filed November 16, 2000.

Both the primary interconnect fabric and the backup interconnect fabric
10 are implemented together in the network. Figure 12 shows the primary and backup interconnect fabrics of the example implemented together in the network. As shown in Figure 12, the primary interconnect fabric connects the first set of ports of each node while the backup interconnect fabric connects the second set of ports of each node.

15 Figure 13 shows a system having a fabric design tool 400 that may employ the method 100 (and the methods 200 and 300) to provide reliability to an interconnect fabric in response to a set of design information 430. The fabric design tool 400 may be implemented in software and/or hardware to perform its functions. The design information 430 in one embodiment includes
20 a list of hosts (source nodes) and devices (terminal nodes) 410, a list of fabric node types 412, a list of link type data 414, a set of flow requirements data 416, a set of port availability data 418, a set of bandwidth data 420, and a set of cost data 422. The design information 430 may be implemented as an information store, such as a file or set of files or a database, etc.

25 The list of hosts and devices 410 may specify the hosts and devices which are to be interconnected by an interconnect fabric design 424.

The list of fabric node types 412 may specify available interconnect devices, such as hubs, routers, switches, etc.

The link type data 414 may specify a list of available communication
30 links that may be employed in the interconnect fabric design 424 and any relevant constraints. There are numerous examples of available

communication links including fiber optic links, fibre channel links, wire-based links, and links such as SCSI as well as wireless links.

The flow requirements data 416 may specify the desired flow requirements for the interconnect fabric design 422. The desired flow requirements may include bandwidth requirements for each pairing of the source and terminal nodes.

The port availability data 418 may specify the number of communication ports available on each source node and each terminal node and each available interconnect device.

The bandwidth data 420 may specify the bandwidth of each host and device port and each type of fabric node and link.

The cost data 422 may specify costs associated with the available communication links and interconnect devices that may be employed in the interconnect fabric design 424. The cost data 422 may also specify the costs of ports for source and terminal nodes and interconnect devices. Other relevant costs may also be indicated.

The interconnect fabric design 424 generated by the fabric design tool 100 includes a list of the physical communication links and interconnect devices and ports, etc. and may include cost data.

The foregoing detailed description of the present invention is provided for the purposes of illustration and is not intended to be exhaustive or to limit the invention to the precise embodiment disclosed. Accordingly, the scope of the present invention is defined by the appended claims.